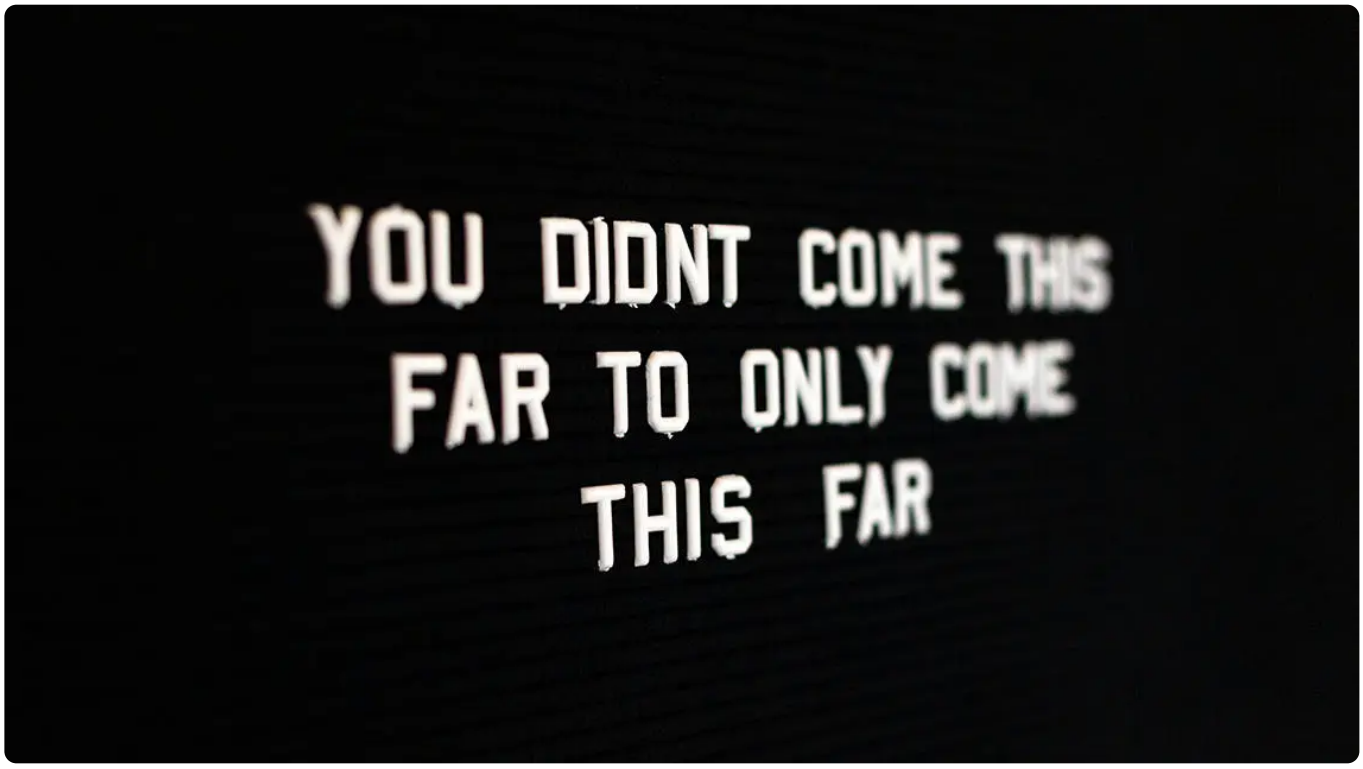
下载APP

即学即练 | 分布式篇：复习卡一键直达

2022-02-19 黄清昊

《算法实战高手课》 课程介绍 >



讲述：代录（老师近期生病）
时长 01:10 大小 1.08M

你好，我是微扰君。

我们已经学完了分布式章，今天就来复习回顾一下这一章的要点内容。既然讲的是工程中的问题，也就没有 LeetCode 练习题了，主要是在工作中多观察多分析多总结。在学习的时候，我们也不要想着只是解决这一个简单的问题，而是要更多地考虑别人解决问题背后的思路。

领资料

有些算法，虽然现在已经有了现成的库类和中间件，供我们开箱即用，但是如果你对它背后的原理更加熟悉，一旦在业务开发中碰到类似的问题，你才能比别人更快地想到有效解决方案，这也是我认为学习这些算法最重要的意义。

分布式篇

今天来复习 & 练习专栏的第四章，**分布式篇**。在这个模块里，我们一起学习了 MapReduce 、PageRank、Raft、UUID、一致性哈希。你可以借助整理好的要点卡片，回到相应章节，有针对性地复习。

复习要点卡

点击要点卡，直达你最需要复习的那篇。

比命令式编程，函数式编程的接口更加声明式，能写出更清晰的代码，且没有状态也能减少 debug 负担。

PageRank 要点卡

- 谷歌对网页检索结果做质量排序的早期策略，本质是图计算的迭代算法，基于图，建立了一种对网页影响力的衡量权重，对搜索结果做排名；

- 对网页上所有页面的链接关系进行构图，迭代地把每个页面的影响力均匀分发

给所有其链向的网页。基于马尔可夫平稳条件，经过若干轮次的迭代，整个图上所有网页的权重最终会收敛到一个稳定的值，此时可以停止计算；

- 为解决 Spider Traps，引入跳转因子，让每个页面增加带有很小权重的指向所有其他节点的边，避免部分节点只有指向自己的出链接，也可以增加权重的平滑性。

Raft 要点卡

- 解决分布式共识问题（一致性问题）的算法，将复杂问题拆分为多个独立步骤和引入更强的约束的方式，实现比 Paxos

更简单；

- 解决基于日志实现的复制状态机中的一致性问题，希望多个集群实例能看起来像整体一样一致地工作，日志一旦被提交，只要多数节点在线，整个系统就能一致地保证此前提交的日志不会丢失；

- 分三步：领导选举、日志复制和安全性。引入任期和日志索引，通过超半数选票，保证每个任期只会出现一个领导人，通过一致性检查，保证每个日志一旦被某任提交，之后任期一定会包含。

UUID 要点卡

- 全局唯一 ID，生成方式可分为两类：引入相当于协调者的第三方组件、基于协议和约定在各个业务节点独立生成；

- 第一类，包括借助数据库的 auto_generate 能力直接生成，以及引入单点借助数据库批量生成慢慢消费，可水平扩展，获得更高的吞吐量，类似批量读写；

- 第二类，UUID 和 Snowflake 的开源实现。UUID 比较长，可直接随机方式生成，冲突概率低；Snowflake 更常见，结合时间戳和序号，通过传入机器标识作为命名，规范可以保证命名出来的不同的 ID 彼此独立，命名相同的 ID 按时序递增。

一致性哈希 要点卡

- 负载均衡问题，集群上多个节点同时对外服务，需要让访问者合理地访问其中的某个节点，同时使得每个节点的负载尽量接近；

- 既可以通过引入一个负载均衡服务实现，也可以直接基于规则和算法在客户端侧实现，常见算法包括：Round-Robin 轮询算法、随机算法、最短响应优先、最少活跃优先、一致性哈希等；

- 一致性哈希算法，适用请求和服务有状态的场景，同一个 key 或者客户端可以在多次请求时访问到同一个节点，将多个节点和请求都映射到一个哈希环上，顺时针访问环上和请求最接近的节点。

复习的过程中，如果有什么困惑，获得了什么收获，都欢迎你留言讨论，也欢迎你把复习卡分享给你的朋友。

学习愉快，我们工程实战篇见～

分享给需要的人，Ta购买本课程，你将得 20 元

生成海报并分享

赞 0 提建议

© 版权归极客邦科技所有，未经许可不得传播售卖。 页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 25 | 一致性哈希：如何在集群上合理分配流量？

下一篇 26 | B+ Tree：PostgreSQL 的索引是如何建立的？

学习推荐

精选留言

写留言

> 由作者筛选后的优质留言将会公开显示，欢迎踊跃留言。